Title
A Commentary On the Use of Sequence Data for Phylogeny Reconstruction - Reply

Journal
Molecular Phylogenetics and Evolution, 4(3)

ISSN
1055-7903

Authors
Avise, JC
Nelson, WS

Publication Date
1995-09-01

DOI
10.1006/mpev.1995.1031

Reply

To the Editor:

Hackett and co-writers are entirely correct that two mtDNA sequences (from a "Jabiru Stork" and a Hoatzin) are in error in our published mtDNA work on numerous avian species. The first of these errors involved a sample misidentification and was indeed egregious. Following discovery of this mistake nearly 1 year ago (shortly after the paper appeared in print), we have made every reasonable effort both to widely publicize and to rectify the problem. This effort involved: (a) immediate removal of the misidentified "Jabiru" sequence from GenBank; (b) complete resequencing of the species and placement of the corrected and independently verified sequence in GenBank (Accession No. U19611); (c) from the near outset, attachment to mailed reprints of a prominent note alerting readers to the error; and (d) publication of a formal correction [*Proc. Natl. Acad. Sci. USA* **92:** 3076 (1995)]. An excerpt from the latter is as follows: "The mtDNA sequence from the sample denoted as the 'Jabiru Stork' is incorrect, the mistake apparently stemming from a sample mix-up, mislabel, or PCR error. Reexamination of *bona fide* Jabiru Stork mtDNA sequences by our laboratory (and independently confirmed elsewhere) now places this species closest phylogenetically to the Wood Stork and Marabou Stork (among the species assayed). We wish to

alert readers to this change and to retract all conclusions regarding the Jabiru Stork from the original paper . . . ." We appreciate the diligent efforts of Hackett *et al.* in once again advertising this issue and thereby sharing our concern that the mistake not be perpetuated in the literature.

The second sequencing difficulty can be traced almost entirely to one specimen ("a") of the Hoatzin, where errors in data transcription were discovered in the last 350 bp (among 961 bp assayed) of the published cytochrome b sequence. We subsequently reanalyzed this specimen as well with corrections deposited in GenBank (U09257). The new sequence for Hoatzin "a" resembles much more closely those of the other conspecifics (maximum sequence divergence $P = 0.005$), so the conundrum of unusually high intraspecific variation (a topic peripheral to the original study) disappears. Fortunately, sequences from the other Hoatzin specimens were essentially correct as published, so the higher level phylogenetic issues remain largely unaffected and the exact phylogenetic position of the Hoatzin vis-à-vis other avian groups remains uncertain from our data.

Criticisms under the next two headings of the Hackett *et al.* commentary also have some validity. True, we might have included locality information for conspecific samples (however, this is neither customary nor invariably encouraged by editors when, as in these studies, the focus is on higher level systematics). True, we might well have assayed more species and included more or better choices as outgroups (such criticisms apply with varying force to all phylogenetic treatments). As additional DNA sequences become available (about 130 avian cytochrome b sequences of at least 950 bp are now in GenBank and the list is growing rapidly), there will be nothing to prevent Hackett *et al.* or others from reexamining the Hoatzin phylogenetic enigma, using whatever outgroup(s) within or outside the class Aves they deem most appropriate. When this is accomplished, something constructive and more definitive may well have been achieved. We only suggest that in conducting these further investigations, analyses not focus myopically on a single algorithm or philosophy of phylogeny reconstruction. This brings us to the final heading of the Hackett *et al.* commentary, which presents the crux of their position on phylogeny estimation. Here, we seem to have a genuine difference in philosophical perspective.

Our approach, which we suggest is appropriately conservative in most contexts, is to apply a variety of conceptually and operationally distinct phylogenetic algorithms to a given molecular data set as coded in various formats. In these studies, for example, we employed three distinct phylogenetic algorithms (maximum parsimony, neighbor-joining, and UPGMA distance analyses) to each of four data partitions (overall nucleotide sequences, transversions only, first and second positions of codons only, and inferred amino acid sequences), for a total of 12 treatments overall. We and no one else can justifiably claim to know which treatment is unequivocally "best" in a given situation (see below), so it seems reasonable to us to compare results of several analyses and to place greater emphasis and confidence in phylogenetic outcomes shared by multiple methods (we are not married to particular ones) than in those which idiosyncratically appear in only one or a few. In our experience, this approach has been quite sobering in the sense of demonstrating the sensitivity of some phylogenetic outcomes to varying analyses (even of one-and-the-same data set). The alternative approach, far more prevalent in the empirical literature, is to present results from a single phylogenetic treatment as if that were the sole gauge of truth.

Contrary to statements by Hackett *et al.*, our position is not equivalent to a claim that all phylogenetic algorithms are equally proficient in recovering phylogeny under all conditions. As judged against computer-simulated evolutionary trees (e.g., Fiala and Sokal, 1985; Sourdis and Krimbas, 1987; Saitou and Imanishi, 1989; Jin and Nei, 1990, 1991) or those generated experimentally in the laboratory (e.g., Hillis *et al.*, 1992; Bull *et al.*, 1993), various phylogenetic algorithms are indeed known to perform differentially in particular circumstances. For example, UPGMA tends to perform poorly when evolutionary rates across branches are heterogeneous (Swofford and Olsen, 1990), as do parsimony analyses particularly when branches show pronounced length differences (the "long-branch-attraction" phenomenon) (DeBry, 1992; Huelsenbeck and Hillis, 1993; Kuhner and Felsenstein, 1994). When evolutionary rates are unequal among nucleotide sites, all phylogenetic methods encounter difficulties, whereas when rate variations among branches and among sites are low, all phylogenetic methods tend to perform reasonably well by most evaluative criteria (Kuhner and Felsenstein, 1994).

The results of many such simulation studies (reviews in Felsenstein, 1988; Nei, 1991; Hillis, 1995) indicate that an algorithm's performance can also vary as a function of the aspect of tree structure examined (topology versus branch length), the size and composition of the data set (e.g., Sourdis and Nei, 1988), and the particular measure of performance monitored (e.g., consistency, efficiency, robustness, discriminating ability, falsifiability, or others). In real-life situations, precise knowledge of DNA evolutionary rates and character-transforming processes is seldom available, so the goodness-of-match to particular assumptions underlying each phylogenetic method usually remains uncertain. Furthermore, alternative phylogenetic methods can differ greatly in other important performance criteria, including computational speed, feasibility of use with large numbers of taxa, transparency of assumptions, avoidance of ambiguity in rooting, and facility in permitting cross-comparisons among studies. Each

method has different strengths and weaknesses that too often have been neglected in a literature focused almost exclusively on recovery of small tree branching topologies alone.

Another important point, often lost or misrepresented by advocates of particular phylogenetic strategies (including Hackett *et al.*), is that even when one phylogenetic algorithm has proven by some criterion to have "outperformed" others in a given condition-dependent computer simulation or experimental test, the *differences* in performance are often relatively minor and/or confined to a subset of the simulated evolutionary conditions (Fiala and Sokal, 1985; Hillis *et al.*, 1992; Kim, 1993). In one illustrative recent example, Huelsenbeck (1995) used computer simulations to examine the capacity of 26 different phylogenetic algorithms to recover correct *branching topology* in a *four-taxon* case when *two sets of branches were allowed to vary in length.* Although "results indicate that most methods perform well (i.e., estimate the correct tree ≥95% of the time) over a large portion of the four-taxon parameter space" (Huelsenbeck, 1995), the UPGMA and Lake's invariants methods were singled out for criticism on the grounds that they failed to consistently recover the correct branching order when the following two conditions simultaneously held: the internal nodes were very close, and the branch lengths to one pair of nonsister taxa were very short in comparison to those of the other pair. In real-life situations, however, we personally would be reluctant to draw firm conclusions about phylogenetic branching order at the organismal level when such conditions hold (regardless of what any single "best" phylogenetic algorithm might otherwise imply), particularly when sequence data from only one or a few genes are available (see below). Instead, our conservative preference would be merely to conclude that the nodes in question were closely spaced (as would likely be well evidenced by short branches in distance-based or other analyses and by inconsistencies in estimated branching topologies across phylogenetic procedures).

Thus, in our opinion, none of the distance-based or parsimony methods commonly employed in the literature warrants the blanket condemnation or applause, respectively, implied by the Hackett *et al.* commentary. Our interpretation of the 30-year-long debate over "best" method for phylogeny reconstruction (partially reviewed in Swofford and Olsen, 1990; Miyamoto and Cracraft, 1991) is in agreement with Felsenstein's (1982) sentiment: "It is essential that we not adopt a single method [of phylogenetic inference] as a universal panacea."

The comparative approach which we advocate helps avoid the pitfall of making any final phylogenetic judgments unduly algorithm- or philosophy-laden. This conservative approach is all the more important when (as is usually the case) an organismal phylogeny is to be inferred from a single gene tree (such as that provided by mtDNA) because true discrepancies between a gene tree and a species tree can also arise from allelic lineage-sorting processes (as well as allelic introgression) across closely spaced nodes, even in the total absence of deficiencies in phylogenetic algorithm performance *per se* (e.g., Neigel and Avise, 1986; Pamilo and Nei, 1988; review in Avise, 1994). We would be surprised if any conventionally employed phylogenetic algorithm (as applied to large data sets with an appropriate window of resolution) consistently failed to identify well-separated organismal clades upon which one could place great confidence. Conversely, we doubt that firm conclusions about organismal phylogeny are normally justified from single-gene nucleotide sequences when putative clades in a gene tree remain idiosyncratic to one or a few philosophically allied phylogenetic algorithms.

In summary, we stand by our statement that was explicitly quoted and criticized by Hackett *et al.*: ". . . outcomes robust to alternative analyses and data bases should be less controversial than those which are strongly analysis- or data-dependent." Indeed, this is a widely held and venerable notion in phylogenetics and one which itself has received recent empirical support from large computer-simulation studies (Kim, 1993). Our preference is that when a putative clade fails to appear consistently across several alternative phylogenetic procedures as applied to a given sequence data set, the more fruitful approach will be to gather and analyze additional data, including those from unlinked DNA regions, rather than to proclaim what may be a fragile truth from one preferred method of phylogeny estimation.

REFERENCES

Avise, J. C. (1994). "Molecular Markers, Natural History and Evolution." Chapman & Hall, New York.

Bull, J. J., Cunningham, C. W., Molineux, I. J., Badgett, M. R., and Hillis, D. M. (1993). Experimental molecular evolution of bacteriophage T7. *Evolution* **47:** 993–1007.

DeBry, R. W. (1992). The consistency of several phylogeny-inference methods under varying evolutionary rates. *Mol. Biol. Evol.* **9:** 537–551.

Felsenstein, J. (1982). Numerical methods for inferring evolutionary trees. *Quart. Rev. Biol.* **57:** 379–404.

Felsenstein, J. (1988). Phylogenies from molecular sequences: Inference and reliability. *Annu. Rev. Genet.* **22:** 521–566.

Fiala, K. L., and Sokal, R. R. (1985). Factors determining the accuracy of cladogram estimation: Evaluation using computer simulation. *Evolution* **39:** 609–622.

Hillis, D. M. (1995). Approaches for assessing phylogenetic accuracy. *Syst. Biol.* **44:** 3–16.

Hillis, D. M., Bull, J. J., White, M. E., Badgett, M. R., and Molineux, I. J. (1992). Experimental phylogenetics: Generation of a known phylogeny. *Science* **255:** 589–592.

Huelsenbeck, J. P. (1995). Performance of phylogenetic methods in simulation. *Syst. Biol.* **44:** 17–48.

Huelsenbeck, J. P., and Hillis, D. (1993). Success of phylogenetic methods in the four-taxon case. *Syst. Biol.* **42:** 247–264.

Jin, L., and Nei, M. (1990). Limitations of the evolutionary parsimony method of phylogenetic analysis. *Mol. Biol. Evol.* **7:** 82–102.

Jin, L., and Nei, M. (1991). Relative efficiencies of the maximum-parsimony and distance-matrix methods of phylogeny construction for restriction data. *Mol. Biol. Evol.* **8:** 356–365.

Kim, J. (1993). Improving the accuracy of phylogenetic estimation by combining different methods. *Syst. Biol.* **42:** 331–340.

Kuhner, M. K., and Felsenstein, J. (1994). A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* **11:** 459–468.

Miyamoto, M. M., and Cracraft, J. (Eds.) (1991). "Phylogenetic Analysis of DNA Sequences." Oxford Univ. Press, New York.

Nei, M. (1991). Relative efficiencies of different tree-making methods for molecular data. *In* "Phylogenetic Analysis of DNA Sequences" (M. M. Miyamoto and J. Cracraft, Eds.), pp. 90–128, Oxford Univ. Press, New York.

Neigel, J. E., and Avise, J. C. (1986). Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. *In* "Evolutionary Processes and Theory" (E. Nevo and S. Karlin, Eds.), pp. 515–534, Academic Press, New York.

Pamilo, P., and Nei, M. (1988). Relationships between gene trees and species trees. *Mol. Biol. Evol.* **5:** 568–583.

Saitou, N., and Imanishi, T. (1989). Relative efficiencies of the Fitch–Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. *Mol. Biol. Evol.* **6:** 514–525.

Sourdis, J., and Krimbas, C. (1987). Accuracy of phylogenetic trees estimated from DNA sequence data. *Mol. Biol. Evol.* **4:** 159–166.

Sourdis, J., and Nei, M. (1988). Relative efficiencies of the maximum parsimony and distance–matrix methods in obtaining the correct phylogenetic tree. *Mol. Biol. Evol.* **5:** 298–311.

Swofford, D. L., and Olsen, G. J. (1990). Phylogeny reconstruction. *In* "Molecular Systematics" (D. M. Hillis and C. Moritz, Eds.), pp. 411–501, Sinauer, Sunderland, MA.

JOHN C. AVISE
WILLIAM S. NELSON

Department of Genetics
University of Georgia
Athens, Georgia 30602